

METHOD FOR ASSEMBLY OF MULTIPLE DNA FRAGMENTS

*Cross -Reference to Related Applications*

[0001] The present application is a continuation of International Application No. PCT/IL98/00096, filed on February 26, 1998, the disclosure of which is incorporated by reference herein.

*Field of The Invention*

[0002] The present invention relates to the field of genetic engineering. More particularly, the invention relates to a novel method for assembling two or more DNA fragments with high efficiency and yield.

*Background of the Invention*

[0003] Building DNA constructs is the basis of genetic engineering. Manipulating recombinant DNA and its incorporation into vectors is common knowledge in the art. For example, fragments obtained by PCR techniques, may be incorporated into plasmids for cloning purposes. Generating specific DNA constructs by joining together different DNA segments, each encoding specific properties and/or possessing specific functions, is also known in the art. The state of the art has been published in a very large number of books, articles, patent applications, patents and the like, and is usually readily available and known to all of skill in the art. For example, a comprehensive account of DNA cloning procedures is provided in the three volume text by Sambrook et al. (1989) entitled "Molecular Cloning-a Laboratory Manual" 2nd edition, Cold Spring Harbor Laboratory Press. This extensive account of the prior art techniques for the combination of DNA fragments, cloning and expression thereof, is included herein by reference, in its entirety.

[0004] According to the present state of the art, the DNA fragments comprising the plasmid to be constructed are cleaved by restriction enzymes from larger entities. Next, they are connected utilizing the enzyme Ligase, which forms phosphodiester bonds between DNA fragments. Restriction enzymes cleave DNA leaving blunt or staggered ends (protruding short single stranded termini) also called overhangs. The efficiency of the ligation of blunt ended fragments is very low. Using staggered ended fragments elevates the efficiency of the ligation process, because the two fragments are being held together by the hydrogen-bonds between the complementary overhangs. Because the overhangs produced by most of the restriction enzymes are very short (2-4 bases), the connections between the complementary overhangs are weak and not very specific. Thus, when constructing a plasmid from restriction enzyme

enzyme produced fragments, the yield is very low and there are many illegitimate ligations, leading to undesired products. It is therefore necessary to amplify the product usually by transfecting cells of choice such as bacterial cells, and then to identify and isolate colonies containing the DNA of choice from the ones containing undesired products. Only then can another DNA fragment be added by repetition of the whole procedure. As a result of this inefficient processes the building of complex DNA constructs is a laborious, time consuming and expensive process, in which each step necessitates the successful completion of the former. The construction of complex molecules may take several weeks to several months. Sometimes the completion of such constructs is not achieved at all. Another major drawback of the above mentioned method is that the availability of restriction site does not always coincide the construction demands. One way to overcome the above mentioned problems is to use short artificial DNA molecules called "linkers". This however further complicates the construction process, reduces the yield, increases the percentage of wrong constructs and sometimes add undesired foreign sequences.

**[0005]** Several methods suggest connecting DNA fragments by producing longer complementary overhangs (8-14 nucleotides). These connections are more stable than the ones produced by restriction enzymes. In fact, they are stable enough to render the ligation step unnecessary and the bacterial cells may be transformed right after the fragments are connected. The phosphodiester bonds are generated later, by the endogenous bacterial ligation machinery.

**[0006]** One prior method uses single strand extensions that are created by adding nucleotides at the 3' end of a DNA strand in a template-independent fashion (Roychoudhury, R. Gene Amplif. Anal. 2:41-83, 1981). The enzyme used in this method, Terminal Transferase, incorporates nucleotides at the end of a double-stranded DNA fragment, thus creating a single-stranded tail. Since the enzyme uses the nucleotides randomly, the only way to ensure that the single-stranded tail will be complementary to a corresponding overhang created on a second DNA molecule, is to provide for each extension only one of the four nucleotides. The overhangs created with this method must therefore be homopolymeric, so that only four types of overhangs can be used, corresponding to the residues dA, dC, dG or dT. Since the overhangs created on both termini of a DNA fragment must be identical, cloning with this method is directionless and can only involve two fragments that are connected to each other at both ends, forming a circular molecule. Furthermore, the length of the overhangs cannot be specifically controlled. Finally, the method necessarily introduces an

unwanted stretch of nucleotides into the final construct, the length of which cannot be determined exactly, making the method unsuitable for the purpose of cloning into expression vectors where the reading frame must be preserved.

**[0007]** According to another method (the commercial product "PCR-Direct™", manufactured by CLONTECH, Inc., USA) the overhangs are generated utilizing the exonuclease activity of a DNA polymerase.

**[0008]** U.S. Patent No. 5,137,814 describes another method in which the overhang is generated by providing at least one dU residue instead of dT, close to the terminus of the fragment. The position of the dU residues determines the length of the overhang. This method involves an a-purination of the Uracil bases. The a-purinated residues no longer have hydrogen-bond connections with their complementary bases on the opposite strand. Moreover, they destabilize the hydrogen bonds of their neighboring bases as well. The resulting 3' protruding termini may connect to a complementary single strand sequence. A commercial product using this method is the CloneAmp® pAMP1 System (manufactured by Life Technologies, Inc., USA). In the aforesaid method a 12-base overhang is used.

**[0009]** The use of overhangs longer than 4 nucleotides for the purpose of fragment cloning is also described in several other publications. Rashtchian et al. (Anal. Biochem. 206, p. 91-97, 1992), describe 12 nucleotide overhangs generated by Uracil DNA Glycosylase (UDG) to achieve high-efficiency cloning of single inserts into a vector. Kuijper et al. (Gene 112, p. 147-155, 1992) and Aslanidis et al. (PCR Methods Appl. 4:172-177, 1994), describe a cloning method wherein T4 polymerase is used together with a predetermined dNTP to generate overhangs of a certain length in PCR products. This method requires a specific sequence to be present in the PCR primer. Hsiao et al. (NAR 21, p. 5528-5529, 1993) and Yang et al. (NAR 21, 1889-1893, 1993) disclose generation of overhangs by the exonucleolytic activity of Exonuclease III (Exo III) or of T4 polymerase. Overhangs of 12 (Aslanidis), 8 (Yang) and 10-14 (Hsiao) nucleotides are disclosed.

**[0010]** The probability of producing a joint molecule composed from three molecules is the probability of joining of the first two molecules, multiplied by the probability of joining of the second and third molecules. If the joining of any two molecule is a rare event, due to the relatively short overhang, the joining of three or more molecules becomes practically useless for cloning purposes, because of the resulting low efficiency.

**[0011]** It is therefore clear that the present state of the art is not satisfactory and there is a need for an improved method by which several DNA fragments can be effectively joined

together in a directional, predetermined way, in one single-step. Such a method will remove the limitations imposed on genetic engineering by the aforementioned methods, and will thus revolutionize the way in which genetic engineering is performed.

#### *SUMMARY OF THE INVENTION*

**[0012]** In one aspect, the invention is directed to a method for assembling two or more DNA fragments with high efficiency, comprising:

a) providing, for each DNA fragment, at least one protruding terminus, or "overhang", capable of hydrogen bonding to a complementary sequence on at least one strand of a second DNA fragment, said overhang having at least 15 bases; and

b) mixing two or more said DNA fragments under conditions suitable to promote joining thereof.

**[0013]** The method of the invention is based on the very surprising finding that increasing the number of bases in complementary overhang from 12 to at least 15, permits to reduce the ratio of the reagents from about 1:100 - 300 to about 1:1. At the same time, this permits to join a plurality of fragments (3 or more), because of the high efficiency of the process.

**[0014]** According to one embodiment of the invention three or more fragments are joined by the method of the invention, and the molar ratio of each DNA pairs is about 1:1 to 1:50.

**[0015]** The present invention provides a DNA fragment comprising an overhang of at least 15 nucleotides or an end portion suitable to be converted into such an overhang. The invention further provides said DNA fragment, for use in the above method of the invention.

**[0016]** According to another preferred embodiment of the invention, the number of bases in the overhang is between 20 and 30. While, as stated above, the improvement of the process efficacy is very dramatic when passing from 12 to 15 bases, there is still a further steep improvement in the efficacy. when increasing the number of bases to about 20. This increase reaches a plateau above 20 bases. Taking into account that, when joining a number of fragments, the overall efficiency of the process decreases, it is of course desirable to employ the lowest number of bases in the overhang which still provides for increased joining efficiency. As stated, in most cases this optimal length will be in the neighborhood of 20 bases.

**[0017]** As will be appreciated by the skilled person, this is most surprising because such a dramatic increase in binding ability of fragments (from almost no binding when

overhangs are 12 bases long to almost 100% binding when the overhangs are 21 bases long) was unexpected. Although *some* improvement could have been expected by increasing the length of the overhang, such a dramatic improvement (two orders of magnitude) was entirely surprising. The art, in fact, has continued to use stepwise cloning techniques, with all the inherent disadvantages, and has not attempted in practice to use overhangs longer than 12 bases.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**[0018]** Figs. 1-6 illustrate plasmids produced in Example 3. In the Figs. 1-6, the dark, cross-hatched bar depicts the fragment amplified from pBR322 which carries the Amp<sup>r</sup> gene and the origin of replication. The empty bar depicts the fragment PCR amplified from pBR322 which carries the Tet<sup>r</sup> gene. The hatched bar depicts the PCR amplified fragment from pACYC184 which carries the Cm<sup>r</sup> gene. The arrows mark primers, where the arrow head is the 3' end of the primer, and the wavy tails mark the various repetitive sequences at the 5' end of the primers. The primer's number is given above or below the arrow. The black area denotes repetitive sequences situated at the connection between two fragments. Fig. 12 illustrates the plasmid produced in example 4. The empty, cross-hatched and hatched bars have the above meanings. The arrows mark the primers which are used for the seamless junction of the fragments.

**[0019]** Fig. 1 is plasmid pGS101. There is an overlap of 12bp between every two fragments which are joined together. The plasmid is assembled from three fragments.

**[0020]** Fig. 2 is plasmid pGS102. There is an overlap of 12bp between every two fragments which are joined together. The plasmid is assembled from four fragments.

**[0021]** Fig. 3 is plasmid pGS103. There is an overlap of 12bp between every two fragments which are joined together. The plasmid is assembled from four fragments.

**[0022]** Fig. 4 is plasmid pGS201. There is an overlap of 20-21bp between every two fragments which are joined together. The plasmid is assembled from three fragments.

**[0023]** Fig. 5 is plasmid pGS202. There is an overlap of 20-21bp between every two fragments which are joined together. The plasmid is assembled from four fragments.

**[0024]** Fig. 6 is plasmid pGS203. There is an overlap of 20-21bp between every two fragments which are joined together. The plasmid is assembled from four fragments.

**[0025]** Fig. 7 shows the results obtained by an Agarose gel run, to check fusion ability of identical fragments with different overhang lengths.

**[0026]** Fig. 8 shows the colonies that grew on plates following transformation with constructs illustrated in Figures 1-6. The results of constructs made with overhangs of 12 bases are shown on the left and those with overhangs of 20-21 bases are shown on the right. Figs. 8 A, B, C, D, E, F, and G correspond to plasmids pGS101, pGS201, pGS102, pGS202, pGS103 and pGS203, respectively. The plating was carried out on selective plates that would minimize background noise as explained in the text.

**[0027]** Fig. 9 illustrates the constructs made in Example 3, along with the number of transformants obtained;

**[0028]** Fig. 10 are the results of two gel runs of the products obtained in Example 3;

**[0029]** Fig. 11 shows the results of gel runs of the products obtained in Example 2;

**[0030]** Fig. 12 shows a construct made by joining 3 different fragments using exonuclease III-created overhangs in the experiment of example 4;

**[0031]** Fig. 13 shows a construct made by joining 8 different fragments in the experiment of example 5.

## DETAILED DESCRIPTION OF THE INVENTION

### General Procedures

#### Overview

**[0032]** The present invention is based on the surprising discovery that by increasing the length of complementary overhangs of two DNA fragments from 12 bases, as in the prior art, to 15 bases or more, the tendency of the fragments to be connected to each other rises, steeply increasing, from close to 0% when 12 base overhangs are used, to almost 100% when around 20 bases are employed. When long complementary overhangs of 15 bases or more are used, a large number of DNA fragments can be assembled in a single step in a predetermined, directional manner. In addition, when connections between just two fragments are made, the yield of cloned products is much higher than that of the prior art, when operating according to the invention.

**[0033]** A number of procedures can be used in order to create these overhangs. In the examples detailed below, three different procedures were used.

#### The first procedure

**[0034]** The first procedure is used in examples 1-3. Examples 1 and 2 illustrate the difference in the connecting ability of fragments with complementary overhangs of different lengths *in-vitro* by subjecting the connected fragments to agarose gel analysis. In the third example, successful cloning of a plasmid assembled from five fragments using the procedure

was demonstrated. The fragments were connected to each other *in-vitro*, and then transformed into bacterial cells.

**[0035]** The procedure is as follows:

**[0036]** The fragments which are to be attached to each other in a directional fashion are preferably prepared by utilizing the well-established Polymerase Chain Reaction (PCR). This provides suitable amounts of the fragments and is especially preferred when the fragments to be used are originally obtained or available only in small amounts, for example from restriction-endonuclease treated plasmid DNA, genomic DNA, or cDNA libraries in which the desired fragments are present in small amounts. In this way, the original fragments are greatly amplified and by virtue of the use of pre-selected specific primers, the ends of the PCR-amplified fragments will contain the desired pre-selected complementary sequences.

**[0037]** The primers are designed such that they consist of two regions: A 3' region that is complementary to the template and a 5' region that contains a sequence complementary to a sequence at the 5' end of the fragment it is designed to join. In this 5' region, several dT residues are substituted by dU residues. During PCR, the thermostable DNA Polymerase does not distinguish between the dT and dU residues and incorporates dA opposite both of these residues in the newly synthesized strand..

**[0038]** Following PCR, the fragments are mixed together and the enzyme Uracyl DNA Glycosylase (UDG) is added to the reaction mixture. UDG specifically attacks the dU bases and a-purinates them. The a-purinated residues no longer have hydrogen bonds with their complementary bases. Furthermore, they destabilize the hydrogen bonds of their neighboring bases as well. As a result, the double stranded structure near the terminus falls apart. Two open strands are created at the end of the segment: 1. A 5' single strand with several a-purinated residues that can no longer form hydrogen bonds with a complementary strand. 2. A 3' single strand with no a-purinated residues that can form such bonds. Although this single strand is not, strictly speaking, an overhang, it is equivalent to it and behaves as such for all practical purposes. The complementary regions of DNA generated in this procedure are made of repetitive sequences. Therefore, it can be used to illustrate that the length of the exposed ends is the key parameter in determining the tendency of fragments to be connected to one other: The difference between the constructs in the various experiments is in the lengths of the repetitive complementary regions and not in the sequence.

**[0039]** Two fragments that have such complementary single strands can form hydrogen-bond connections with each other. If such fragments are used to transform bacterial cells, their 5' a-purinated single strands are degraded by cellular enzymes. Furthermore, endogenous ligases convert the joint fragments into a single DNA molecule. Details of the procedure are described in Examples 1-3 below.

*The second procedure*

**[0040]** In Example 4, a second procedure was used. This procedure uses the enzyme Exonuclease III (herein after ExoIII) to generate the overhangs. Other enzymes having an exonuclease activity such as T4 DNA polymerase or T7 DNA polymerase may be used as well. In addition to the above mentioned enzymes which have 3'→5' exonuclease activity, enzymes with 5'→3' exonuclease activity may also be used, depending on the circumstances. Procedures similar to this one are disclosed in the above Hsiao et al. and Yang et al. The specific procedure is elaborated in a copending application of the same applicant herein (filed on the same day as this application, and identified as Attorney Docket 4190/96, Israeli Patent Application No. 120378). The procedure can be applied for the joining of PCR fragments as in examples 1-3 as well as for the joining of a mixture of DNA fragments that are generated by restriction enzyme cleavage together with PCR generated fragments. In example 4 the fragments were produced by PCR to generate a plasmid assembled from three segments.

**[0041]** The procedure is as follows:

**[0042]** The terminal portions of the fragments that are to be joined have to be complementary to each other. This can be achieved by producing the PCR fragments in the following way: The primers of a first DNA fragment that is to be PCR produced and is to be joined to a second and to a third fragment should be synthesized having two regions: 1. A 3' region that is complementary to the template. 2. A 5' region that is complementary to the terminal portion of the fragment that is to be joined to the first fragment. The length of the complementary regions should be 15 nucleotides (nt) or more.

**[0043]** The various DNA fragments to be joined are mixed together and an exonuclease enzyme is added to the mixture. In the case of example 4, ExoIII was used. ExoIII is an exonuclease that catalyzes the stepwise removal of mononucleotides from 3' hydroxyl termini of double stranded DNA. It is added in excess and the reaction is incubated at 6°C to ensure slow digestion. The reaction is allowed to proceed for a time sufficient to



digest the ends of the fragments, exposing 5' overhangs. These overhangs usually consist of a region which is longer than the complementary portion.

**[0044]** Once the reaction is stopped, the temperature is raised to 75°C. The reaction mixture is then cooled slowly. This facilitates the joining of complementary overhangs while minimizing illegitimate connections. Following the cooling, dNTPs and T7 DNA Polymerase are added in order to fill any gaps that may have been formed due to ExoIII overdigestion. At this stage, the mixture is used to transform bacterial cells. Further elaboration of the procedure is given in Example 4 below.

#### *The third procedure*

**[0045]** In example 5 a third procedure was used. This procedure is similar, in many respects to the first one but has the advantage that the junctions between the various fragments require no addition of linker nucleotides. An elaborated description of this procedure is disclosed in a copending application (filed on the same day as this application, and identified as Attorney Docket 4149/96, Israeli Patent Application No. 120377). As will be demonstrated below, a single dU residue near the termini of the fragments, along with UDG and the reagent N,N dimethylethylenediamine were employed in this procedure to generate seamless connections between fragments. In example 5, the cloning of a plasmid assembled from eight separate segments is demonstrated. To date, this is the only disclosure of an assembly of a plasmid from so many segments in a single step.

**[0046]** The procedure is as follows:

**[0047]** This procedure allows the formation of true 3' overhangs using a single dU residue in each of the primers to be connected. It is based on the ability to chemically create a nick at the 3' end of an a-purinated dU residue.

**[0048]** The fragments are produced by PCR. As in the other examples, the terminal portions of the fragments that are to be joined have to be complementary to each other. As described in example 4, this is achieved by producing the PCR fragments in the following way: The primers of a DNA fragment that is to be PCR produced and is to be connected to two other fragments are synthesized having two regions: 1. A 3' region that is complementary to the template. 2. A 5' region that is complementary to the terminal portion of the fragment that is to be joined to the said fragment. The length of the complementary regions should be 15 nt or more. In each primer, a dT residue is substituted by a dU residue at the junction between the complementary and non complementary regions.

**[0049]** The various DNA fragments to be joined are mixed together and UDG, as well as N,N dimethylethylenediamine are added to the mixture. The enzyme a-purinates the dU base and the chemical generates a nick at the 3' side of the a-purinated residue.

**[0050]** Following the creation of the above mentioned nicks, the temperature is raised to 70°C in order to disconnect the short single-strand stretch 5' of the a-purinated dU residues from the rest of the fragment. Thus, 3' overhangs which constitute the complementary regions are created. After removing the nicked oligonucleotides, the reaction is cooled slowly to allow joining between the complementary overhangs while minimizing illegitimate connections. Following the cooling, the mixtures are used to transform bacterial cells. Further elaboration of the procedure is found in Example 5 below.

**[0051]** The examples to follow are provided for the purpose of illustration only, and it is clear that the invention is not limited to any particular method or procedure for creating the aforesaid overhangs.

### *Example 1*

#### *In Vitro Hybridization of Fragments with Complementary Overhangs of Different Lengths* *Overview*

**[0052]** In order to demonstrate the superiority of connection between DNA fragments having long overhangs, the degree of inter-connection of fragments with different lengths of overhangs was tested. Five pairs of fragments, each containing different length of the complementary region, were produced by PCR. Each pair consisted of: 1). A fragment of about 1400 bp designated Tet, which was PCR amplified from the Tet<sup>r</sup> gene (conferring resistance to tetracycline) of pBR 322. 2). A fragment of about 1700 bp designated Amp, which was PCR amplified from the Amp<sup>r</sup> gene (conferring resistance to ampicillin) of pBR322. One of the two primers that was used to amplify the Tet fragment in each pair contained, in addition to a region which was complementary to the template DNA, a 5' tail with repeats of three bases CUA. Likewise, one of the primers used for Amp amplification contained a 5' tail with repeats of UAG. Note, that CUA and UAG are complementary to one another. In the first pair, the tail consisted of four repeats (12 bases) and the tails of the second, third, fourth and fifth pairs consisted of five, six, seven and ten repeats respectively (15, 18, 21 and 30 bases).

**[0053]** The Amp fragment and the Tet fragment of each pair were mixed, and their 3' overhang exposed utilizing UDG. After an annealing time period the mixtures were subjected to agarose gel analysis (Fig. 7).

[illegible]

[0055] The hybridization rate between pairs of DNA fragments was tested.

[0056] In each of the primers below, and in the following examples, the \* symbol is written below the dU residue in order to emphasize it. All of the primer sequences given herein are written in 5' to 3' direction.

(also designated 31160) - upstream of Tet (PBR322: 1-22)

note: the 5 primers below differ from one another only by the number of CUA repeats at their 5' end.

(also designated 3781) - downstream of Tet (pBR322: 1449-1425)

Primer SEQ. ID NO. 3

(also designated 4082) - downstream of Tet (pBR322: 1449-1425)

Primer SEQ. ID NO. 4

(also designated 4353) - downstream of Tet (pBR322: 1449-1425)

Primer SEQ. ID NO. 5

(also designated 4635) - downstream of Tet (pBR322: 1449-1425)

---21nt (CUA)<sub>x7</sub>--- TGG CCA GGA <sup>25nt</sup> CCC AAC GCT GCC CGA G

Primer SEQ. ID NO. 6

(also designated 5535) - downstream of Tet (pBR322: 1449-1425)

25nt

---30nt (CUA)x10--- TGG CCA GGA CCC AAC GCT GCC CGA G

\*

(Note that primers SEQ. ID NO. 2-6 are homologously complementary to primers SEQ. ID NO. 8-12 respectively at their 5' ends)

Primer SEQ. ID NO. 7 - *Amp sense primer*

(also designated 4142) - downstream of Amp (pBR322 2460-2479)

-----irrelevant tail----- 20nt:

ATTGGTGCCCTTAAACGCCTG-AACGCAGGAAAGAACATGTG

note: the 5 primers below differ from one another only by the number of UAG repeats at their 5' end.

#### *Amp antisense primers*

Primer SEQ. ID NO. 8

(also designated 36107) - upstream of Amp (pBR322: 4136-4159)

24nt

---12nt (UAG)x4---AAG AGT ATG AGT ATT CAA CAT TTC

\*

Primer SEQ. ID NO. 9

(also designated 3993) - upstream of Amp (pBR322: 4136-4159)

24nt

---15nt (UAG)x5---AAG AGT ATG AGT ATT CAA CAT TTC

\*

Primer SEQ. ID NO. 10

(also designated 4270) - upstream of Amp (pBR322: 4136-4159)

24nt

---18nt (UAG)x6---AAG AGT ATG AGT ATT CAA CAT TTC

\*

Primer SEQ. ID NO. 11

(also designated 4560) - upstream of Amp (pBR322: 4136-4159)

24nt

---21nt (UAG)x7---AAG AGT ATG AGT ATT CAA CAT TTC

\*

Primer SEQ. ID NO. 12

(also designated 5434) - upstream of Amp (pBR322: 4136-4159)

24nt

---30nt (UAG)<sub>x</sub>10---AAG AGT ATG AGT ATT CAA CAT TTC

\*

Note that primers SEQ. ID NO. 8-12 are homologously complementary to primers SEQ. ID NO. 2-6 respectively at their 5' ends.

**[0057]** The template for the PCR reaction was pBR 322 DNA. The full sequence and maps of the various regions of this plasmid is well known and can be accessed from GenBank database under accession no. J01749 (pBR322).

**[0058]** The various fragments were produced by PCR, as follows:

*PCR of Tet fragments*

(the number at the right denotes the length of the overhang)

	Tet 12	Tet 15	Tet 18	Tet 21	Tet 30
10µl (0.1 µg/µl)	SEQ ID 2	SEQ ID 3	SEQ ID 4	SEQ ID 5	SEQ ID 6
10µl (0.1 µg/µl)	SEQ ID 1	SEQ ID 1	SEQ ID 1	SEQ ID 1	SEQ ID 1
pBR322 DNA	40 ng	40 ng	40 ng	40 ng	40 ng
180µl PCR mix in each tube					

*PCR of Amp fragments*

(the number at the right denotes the length of the overhang)

	Amp 12	Amp 15	Amp 18	Amp 21	Amp 30
10µl (0.1 µg/µl)	SEQ ID 8	SEQ ID 9	SEQ ID 10	SEQ ID 11	SEQ ID 12
10µl (0.1 µg/µl)	SEQ ID 7	SEQ ID 7	SEQ ID 7	SEQ ID 7	SEQ ID 7
pBR322 DNA	40 ng	40 ng	40 ng	40 ng	40 ng
180µl PCR mix in each tube					

PCR mix:

dNTPs (2.5mM each)	128µl
Buff x1O	160µl
H <sub>2</sub> O	1088µl
Taq DNA Polymerase 5U/µl	6.4µl

*Temperature regime*

94°C 40sec

40°C 2min

72°C 4min

30 cycles

72°C 5min

6°C infinitely

**[0059]** All the required fragments were obtained. The concentration of each fragment was determined by measuring the density of the band created by each DNA fragment on an agarose gel. The volumes used for the following experiment were corrected accordingly.

*B) Checking fusion ability*

**[0060]** Each pair of fragments having the same length of repeats was put in a different Eppendorf tube. The DNA was incubated at 37°C with UDG in order to create the overhangs and allow hybridization (see details below). The products of the reaction were subjected to electrophoresis on an agarose gel at 60V for 60min, stained with ethidium bromide and photographed under U.V. light (Fig. 7). Amp fragments are 1.7kb in length, Tet fragments are 1.4kb, and fragments of Amp+Tet that have hybridized to each other are 3.1 kb in length. The difference in the density between the Amp band, the Tet band and the Amp+Tet band, reflect the ability of the fragments to adhere to each other. Comparisons could thus be made between the adherence abilities of fragment-pairs having different overhang-lengths.

**[0061]** The Amp and Tet fragments (of each pair) were mixed together as follows (The number of the mix is equivalent to the number of the complementary overhang).

Mix	12	15	18	21	30
Amp frag.	3µl	1.5µl	1µl	1µl	1µl
Tet frag.	3µl	2µl	2.5µl	1.5µl	5µl
H <sub>2</sub> O	2µl	4.5µl	4.5µl	5.5µl	2µl
UDG Buff. x10	1µl	1µl	1µl	1µl	1µl

UDG Buff.x10= 200mM Tris HCl pH 8.4, 500mM KCl, 15mM MgCl<sub>2</sub>

**[0062]** The content of each of the tubes was divided into two. In the first tube, 0.5µl UDG was added. In the second, 0.5µl H<sub>2</sub>O was added as a blank control. The tubes were incubated at 37°C for 5 hours. Then, the content of the tubes was transferred into an 0.8% TAE Agarose gel. The gel was run at 60V for 60min, stained with ethidium bromide and photographed under U.V. light. Results are shown in Fig. 7.

OVERHANG	12		15		18		21		30	
Tube No.	1	2	3	4	5	6	7	8	9	10
	UDG H <sub>2</sub> O		UDG H <sub>2</sub> O		UDG H <sub>2</sub> O		UDG H <sub>2</sub> O		UDG H <sub>2</sub> O	
Lane on the gel:	1	2	3	4	5	6	7	8	9	10

[0063] The experiment shows: 1) Almost no fusion is observed when the overhangs are 12bp long. 2) In lane 7, there are more Tet fragments than Amp fragments. Therefore, even though there are almost no free Amp fragments left, some free Tet fragments still exist. 3) Both in lanes 7 and 9, there are no free Amp fragments left. There is an outstanding ability of overhangs of 21 and 30bp to adhere to one another.

[0064] In conclusion, the results of the experiment indicate that:

- 1) Almost no fusion is observed when the overhangs are 12bp long.
- 2) There is a dramatic increase in adherence starting with overhangs 15 bases long.
- 3) There is an outstanding ability of almost 100% of overhangs of 21 and 30 bases to adhere to one another.

### *Example 2*

#### *Multi DNA Fragment Assembly (MDFA) in-vitro: Comparison between fragments with overhangs of 12 and 20-21 bases*

##### *I. Overview:*

[0065] The following experiment was carried out in order to show the stronger affinity and higher stability of connection between fragments with overhangs of 21 bases versus the connection of fragments having overhangs of 12 bases. Two sets of fragments were generated by PCR. In the first set, the length of the complementary regions at each end of the fragments was 12 bp. In the second set, the length was 20-21bp. The complementary regions were repeats of 3 or 5 bases. Therefore, the sequence of the complementary regions in analogous fragments of the two sets can be regarded as identical, the only difference being the overall length of the repeated region. Four different fragments of each set were generated by PCR, mixed together and 3' complementary overhangs were exposed utilizing UDG. After an annealing time period, they were run on agarose gel in order to check the degree of the connection between the fragments (Fig. 11).

##### *II. Primers and PCR conditions:*

[0066] Four unique fragments of each set (one set with 12 bp complementary sequences and the other set with 21bp complementary sequences) were amplified by PCR. The fragments were designated TetA, TetB, Amp, and CmB. TetA was PCR amplified from pBR322 (bases 1-937). TetB was PCR amplified from pBR322 (bases 1067-1449). Amp was PCR amplified from pBR322 (bases 2460-4159). CmB was PCR amplified from pACYC184 (bases 3768-4086).

**[0067]** The fragments were designed to connect with each other in the following order: TetA-TetB-Amp-CmB. The primers were synthesized with 3' ends that would enable making PCR fragments from pBR322 and pACYC184 templates as in the former experiments. Two sets of 8 primers were synthesized. In each set, three junctions with complementary repeats were designed. A junction between the TetA and the TetB fragments, a junction between the TetB and the Amp fragments and a junction between the Amp and the CmB fragments. Each of the repeats in a set was different and was complementary-homologous only to one of the other repeats. The PCR fragments that were produced could therefore be connected to one another via the repeats, to form a linear fusion product.

*Primers of the first set: primers containing repeats or near-repeats of 12 nucleotides with dUs at their 3' ends*

Primer SEQ ID NO. 13 - Tet A sense

(also designated 37101)

upstream of Tet (pBR322: 1-25)

```

-----12nt-----                25nt
AU CAUAU CAUAU TTC TCA TGT TTG ACA GCT TAT CAT C
  *   *   *   *

```

Primer SEQ ID NO. 15 - Tet A antisense

(also designated 37103)

```

-----12nt-----                25nt (pBR322: 913-937)
UA CUAUA CUAUA CAT GCC GGC GAT AAT GGC CTG CTT C
  *   *   *   *

```

Primer SEQ ID NO. 16 - Tet B sense

(also designated 37104)

```

-----12nt-----                25nt (pBR322: 1067-1091)
UAUAG UAUAG UA GTA GAT GAC GAC CAT CAG GGA CAG C
  *   *   *   *

```

Primer SEQ ID NO. 2 Tet B antisense

downstream of Tet (pBR322: 1449-1425)

```

-----12nt-----                25nt
CUA CUA CUA CUA TGG CCA GGA CCC AAC GCT GCC CGA G
  *   *   *   *

```

Primer SEQ ID NO. 8 - Amp sense

upstream of Amp (pBR 322: 4159-4136)

```

-----12 nt UAGx4 -                24 nt
UAG UAG UAG UAG AAG AGT ATG AGT ATT CAA CAT TTC
  *   *   *   *

```

Primer SEQ ID NO. 18 - antisense Amp

(also designated 36108)

downstream of ORI



-----12 nt AUGx4 - pBR322 (2460-2483)  
 AUG AUG AUG AUG AAC GCA GGA AAG AAC ATG TGA GCA  
 \* \* \* \* \*

Primer SEQ ID NO. 17 - Cm B sense  
 (also designated 37106)

-----12nt ----- 25nt (pACYC184: 4086-4062)  
 AU CUAUU CUAUU TAC GGT GAA AAC CTG GCC TAT TTC C  
 \* \* \* \* \*

Primer SEQ ID NO. 14 - Cm B antisense  
 (also designated 3866)

-----12nt CAUx4----- 26nt (pACYC184: 3768-3793)  
 CAU CAU CAU CAU CAG GCG TTT AAG GGC ACC AAT AAC TG  
 \* \* \* \* \*

### PCR REACTIONS

(The number on the right denotes the length of the overhang)

reaction	Tet A12	Tet B12	Amp12	Cm B12
20ml primer 1	SEQ ID 13	SEQ ID 16	SEQ ID 18	SEQ ID 14
20ml primer 2	SEQ ID 15	SEQ ID 2	SEQ ID 8	SEQ ID 17

(Primer concentration: 0.1 µg/µl)

expected

fragment size:	961bp	406bp	1724bp	340bp
mix				

2µl dNTP's (2.5 mM each)	112µl
2.5µl 10xBuff.	140µl
18µl H2O	1008µl
0.1µl Taq (5u/µl)	5.6µl

put 360µl mix in each tube. add primers

add 1µl DNA	pBR322	pBR322	pACYC184	pBR322
-------------	--------	--------	----------	--------

(DNA concentration: 40 ng/µl)

### Temperature regime

94°C 40sec  
 40°C 2min  
 72°C 4min  
 30 cycles  
 72°C 5min  
 6°C infinitely

*Primers of the second set: primers containing repeats of 21 nucleotides with dUs at their 3' ends*

Primer SEQ. ID NO. 19 - TetA sense

(also designated 4554) - upstream of Tet (pBR322: 1-25)

-----20nt CAUAUx4----- 25nt  
CAUAU CAUAU CAUAU CAUAU TTC TCA TGT TTG ACA GCT TAT CAT C  
\* \* \* \* \*

Primer SEQ. ID NO. 21 - TetA antisense

(also designated 4558) - (pBR322: 913-937)

-----20nt CUAUAx4----- 25nt  
CUAUA CUAUA CUAUA CUAUA CAT GCC GGC GAT AAT GGC CTG CTT C  
\* \* \* \* \*

Primer SEQ. ID NO. 22 - TetB sense

(also designated 4559) - (pBR322: 1067-1091)

-----20nt UAUAGx4----- 25nt  
UAUAG UAUAG UAUAG UAUAG GTA GAT GAC GAC CAT CAG GGA CAG C  
\* \* \* \* \*

Primer SEQ. ID NO. 5 - TetB antisense

downstream of Tet (pBR322: 1449-1425)  
-----21nt CUAx7----- 25nt  
CUA CUA CUA CUA CUA CUA CUA TGG CCA GGA CCC AAC GCT GCC CGA G  
\* \* \* \* \*

Primer SEQ. ID NO. 11 - Amp sense

- pBR322 (4159-4136)  
-----21nt UAGx7----- 24nt  
UAG UAG UAG UAG UAG UAG UAG AAG AGT ATG AGT ATT CAA CAT TTC  
\* \* \* \* \*

Primer SEQ. ID NO. 24 Amp antisense

(also designated 4561) - pBR322 coding (2460-2483)  
-----21nt AUGx7----- 24 nt  
AUG AUG AUG AUG AUG AUG AUG AAC GCA GGA AAG AAC ATG TGA GCA  
\* \* \* \* \*

Primer SEQ. ID NO. 23 - Cm B sense

(also designated 4556) - (pACYC184: 4086-4062)  
-----20nt CUAUx4----- 25nt  
CUAAU CUAU CUAU CUAU TAC GGT GAA AAC CTG GCC TAT TTC C  
\* \* \* \* \*

Primer SEQ. ID NO. 20 - CmB antisense

(also designated 4743) - (pACYC184: 3768-3793)  
-----21nt CAUx7----- 26nt  
CAU CAU CAU CAU CAU CAU CAU CAG GCG TTT AAG GGC ACC AAT AAC TG  
\* \* \* \* \*

# **PCR REACTIONS**

reaction	Tet A21	Tet B21	Amp21	Cm B21
20µl primer 1	Seq. ID 19	Seq. ID 22	Seq. ID 11	Seq. ID 20

20µl primer 2 Seq. ID 21 Seq. ID 5 Seq. ID 24 Seq. ID 23  
 (primer concentration: 0.1 µg/µl)  
 expected  
 fragment size: 977bp 423bp 1742bp 357bp  
 mix x128  
 2µl dNTP's (2.5 mM each) 256µl  
 2.5µl 10xBuff. 320µl  
 18µl H2O 2304µl  
 0.1µl Taq 5U/µl 12.8µl  
 (dNTP concentration: 2.5 mM each)  
 put 360ml mix in each tube. add primers  
 add 1µl DNA pBR322 pACYC184 pBR322  
 (DNA concentration: 40 ng/µl)

#### *Temperature regime*

94°C 40sec  
 40°C 2min  
 72°C 4min  
 30 cycles  
 72°C 5min  
 6°C infinitely

### *III. Fusion of the fragments*

**[0068]** After the amplification, the fragments were mixed and 3' complementary overhangs were exposed utilizing UDG. Fusions of either 2 fragments, 3 fragments or 4 fragments of the two sets (i.e., of fragments that were created with primers of the two above-described sets) were performed. After the fusion the reactions were run on a 0.8% agarose gel in 1x TAE stained with ethidium bromide and photographed under U.V. light (Fig. 11).

**[0069]** In order to demonstrate the size of the various fragments, they were run in separate lanes (TetA, TetB, Amp, and CmB, in lanes 1, 4, 7, and 10 respectively). Only fragments of the first set were run, since those of the second set have identical lengths. The various mixtures that included the fused fragments were also run.

**[0070]** Reactions and lanes no. on the gel:

	1.8µl					
1.	Tet A12					
	1.8µl	1.8µl	13.6µl	2µl		0.8µl
2.	Tet A21 Tet B21 H2O	UDG Buffx10		UDG		
	1.8µl	1.8µl	13.6µl	2µl		0.8µl
3.	Tet A12 Tet B12 H2O	UDG Buffx10		UDG		
	1.8µl					

4.	Tet B12							
	1.8µl	1.8µl		1.8µl	11.8µl	2µl		0.8µl
5.	TetA21	TetB21	Amp21	H <sub>2</sub> O	UDG Buffx10		UDG	
	1.8µl	1.8µl	1.8µl	1.8µl	11.8µl	2µl		0.8µl
6.	TetA12	TetB12	Amp12	H <sub>2</sub> O	UDG Buffx10		UDG	
	1.8µl							
7.	Amp12							
	1.8µl	1.8µl	1.8µl	1.8µl	10µl	2µl		0.8µl
8.	TetA21	TetB21	Amp21	CmB21	H <sub>2</sub> O	UDG Buffx10	UDG	
	1.8µl	1.8µl	1.8µl	1.8µl	10µl	2µl		0.8µl
9.	TetA12	TetB12	Amp12	CmB12	H <sub>2</sub> O	UDG Buffx10	UDG	
	1.8µl							
10.	CmB12							

#### 11. Molecular Weight Markers

the concentration of all the DNA fragments was 0.09 pmol/µl.

UDG Buff.x10= 200mM Tris HCl pH 8.4, 500mM KCl, 15mM MgCl<sub>2</sub>

**[0071]** The reactions were incubated at 37°C for 5 hours and then run on a gel. The results are shown in Fig. 11. The order of the lanes is shown at the top, and the sizes of the molecular weight markers are indicated at the right side of the photograph.

**[0072]** From the above results, the following conclusions can be drawn:

1. Detectable fusion products are observed when the overhangs are 21 bp long, but not when they are 12bp long.

2. Since the fusion efficiency of 21 bp overhangs is high but less than 100%, products of partial fusions are also observed. In lanes 2, 5 and 8, the band which appears highest on the gel, has the length of the required end product.

3. The results demonstrate that the amount of the required product decreases when the number of fragments required to generate it, increases.

This example clearly demonstrated the advantage of using fragments with long overhangs. No end products were observed when using fragments with 12 nt overhangs. When using fragments with 20-21 nt overhangs, the fusion of 2-3 and 4 fragments is clearly demonstrated.

#### Example 3

*Multi DNA Fragment Assembly (MDFA) in-vivo: Comparison between fragments with overhangs of 12 and 20-21 bases*

**[0073]** The following experiment illustrates the advantage of connecting fragments having 20-21 base overhangs as compared to fragments having 12 base overhangs, *in vivo*. Fragments having either 12 base overhangs or 20-21 base overhangs were produced by PCR and connected, forming circular plasmids. These plasmids were transformed into bacterial cells and the colonies were examined.

**[0074]** Two sets of plasmids were constructed, as depicted in Figs. 1-6. In the first set, fragments containing 12 base overhangs were used (Figs. 1-3). In the second, fragments containing 20-21 base overhangs were used (Figs. 4-6).

**[0075]** Each set contained three plasmids:

1. A plasmid constructed out of the three basic fragments.
  - an Amp<sup>r</sup> fragment (hereinafter Amp) which contained the Ampicillin resistance gene.
  - a Tet<sup>r</sup> fragment (hereinafter Tet) which contained the Tetracycline resistance gene.
  - a Cm<sup>r</sup> fragment (hereinafter Cm) which contained the Chloramphenicol resistance gene.
2. A plasmid constructed out of four fragments; the said Amp and Cm fragments and two fragments designated TetA and TetB, which are two separate regions of the Tet gene.
3. A plasmid constructed out of four fragments; the said Amp and Tet fragments and two fragments designated CmA and CmB, which are two separate regions of the Cm gene.

**[0076]** The plasmids of the first set are equivalent to the plasmids of the second set except for the length of the overhangs produced. The overhangs of the fragments were designed and produced by adding short repeats of either 3 or 5 bases. Each of the repeats in a set, was different and was complementary to one of the other repeats. The PCR fragments that were produced could therefore be connected to one another via the repeats to form a full circular plasmid (see Figs. 1-6). Analogous primers of the two sets were nearly identical, the only difference being in the length of the repeat sequence. In the first set, the overall lengths of the repeats was 12 bp. In the second, the overall length of the repeats was 20-21 bp. In all cases, the dT residues in the repeats have been changed to dU residues. The 5' termini of fragments containing the repeats could therefore be dissociated from the complementary strand by a-purinating the dU residues, exposing a 3' overhang of the required size.

**[0077]** After amplifying the fragments comprising each plasmid they were mixed and their termini exposed utilizing UDG. After an annealing time period each mixture was used to transform bacterial cells, as detailed below.

A) Primers and PCR conditions

**[0078]** Primers were synthesized with 3' ends that would enable generating PCR fragments from pBR322 and pACYC184 templates. The full sequence and maps of the various regions of these plasmids are well known and can be accessed from GenBank database under accession nos. J01749 (pBR322), and X06403 (pACYC184).

**[0079]** Two sets of primers were synthesized:

*First set of primers: primers containing repeats or near-repeats of 12 nucleotides with dUs at their 3' ends*

Seq. ID NO. 13 - Tet A sense

upstream of Tet (pBR322: 1-25)  
 -----12nt-----  
 AU CAUAU CAUAU TTC TCA TGT TTG ACA GCT TAT CAT C  
 \* \* \* \* \*

Seq. ID NO. 15 - Tet A antisense

downstream of Tet (pBR322: 913-937)  
 -----12nt-----  
 UA CUAUA CUAUA CAT GCC GGC GAT AAT GGC CTG CTT C  
 \* \* \* \* \*

Seq. ID NO. 16 - Tet B sense

(pBR322: 1067-1091)  
 -----12nt-----  
 UAUAG UAUAG UA GTA GAT GAC GAC CAT CAG GGA CAG C  
 \* \* \* \* \*

Seq. ID NO. 2 - Tet B antisense

downstream of Tet (pBR322: 1449-1425)  
 -----12nt CUAx4-----  
 CUA CUA CUA CUA TGG CCA GGA CCC AAC GCT GCC CGA G  
 \* \* \* \* \*

Seq. ID NO. 18 - Amp sense

- upstream of Amp (pBR322 2460-2483)  
 -----12 bp AUGx4-----  
 AUG AUG AUG AUG AAC GCA GGA AAG AAC ATG TGA GCA  
 \* \* \* \* \*

Seq. ID NO. 8 - Amp antisense

- downstream of Amp (pBR322: 4159-4136)  
 -----12 bp UAGx4-----  
 UAG UAG UAG UAG AAG AGT ATG AGT ATT CAA CAT TTC  
 \* \* \* \* \*

Seq. ID NO. 17 - Cm B sense

- (pACYC184: 4086-4062)  
 -----12bp-----  
 AU CUAUU CUAUU TAC GGT GAA AAC CTG GCC TAT TTC C  
 \* \* \* \* \*

Seq. ID NO. 14 - Cm B antisense

downstream of Cm (pACYC184: 3768-3793)  
 -----12nt CAUx4-----  
 CAU CAU CAU CAU CAG GCG TTT AAG GGC ACC AAT AAC TG  
 \* \* \* \* \*

Seq. ID NO. 25 - Cm A sense

(also designated 37102) - upstream of Cm (pACYC184: 240-216)  
 -----12nt----- 25nt  
 AUAUG AUAUG AU TCA GGA GCT AAG GAA GCT AAA ATG G  
 \* \* \* \* \*

Seq. ID NO. 26 - Cm A antisense

(also designated 37105) - (pACYC184: 22-46)  
 -----12bnt----- 25nt  
 AUUAG AUUAG AU CAG GCG GGC AAG AAT GTG AAT AAA G  
 \* \* \* \* \*

# *PCR reactions*

reaction	Tet A12	Tet B12	Cm B12	Cm A12
20µl primer 1	SEQ. ID 13	SEQ. ID 16	SEQ. ID 14	SEQ. ID 25
20µl primer 2	SEQ. ID 15	SEQ. ID 2	SEQ. ID 17	SEQ. ID 26

(primer concentration: 0.1 µg/µl)

expected

fragment size:	961bp	406bp	340bp	241bp
reaction	Tet12	Cm12	Amp12	
20µl primer 1	SEQ. ID 13	SEQ. ID 14	SEQ. ID 18	
20µl primer 2	SEQ. ID 2	SEQ. ID 25	SEQ. ID 8	

(primer concentration: 0.1 µg/µl)

expected

fragment size	1473bp	739bp	1724bp
mix	x128		
2µl dNTP's (2.5 mM each)	256µl		
2.5µl 10xBuff.	320µl		
18µl H2O	2304µl		
0.1µl Taq (5U/µl)	12.8µl		

put 360µl mix in each tube. add primers

add 1µl DNA (40ng/µl):

Tet A12	Tet B12	Cm B12	Cm A12
pBR322	pBR322	pACYC184	pACYC84
Tet12	Cm12	Amp12	
pBR322	pACYC184	pBR322	

# *Temperature regime*

94°C 40sec  
 40°C 2min  
 72°C 4min  
 30 cycles  
 72°C 5min

6°C infinitely

*Second set of primers: primers containing repeats of 21 nucleotides with dUs at their 3' ends*

SEQ. ID NO. 19 - TetA sense

- upstream of Tet (pBR322: 1-25)

-----20nt CAUAX4----- 25nt  
CAUUAU CAUUAU CAUUAU CAUUAU TTC TCA TGT TTG ACA GCT TAT CAT C  
\* \* \* \* \*

SEQ. ID NO. 5 TetB antisense

- downstream of Tet (pBR322: 1449-1425)

----- 21nt CUAX7 ----- 25nt  
CUA CUA CUA CUA CUA CUA CUA TGG CCA GGA CCC AAC GCT GCC CGA G  
\* \* \* \* \*

SEQ. ID NO. 21 - TetA antisense

(pBR322: 913-937)

-----20nt CUAUAx4----- 25nt  
CUAUA CUAUA CUAUA CUAUA CAT GCC GGC GAT AAT GGC CTG CTT C  
\* \* \* \* \*

SEQ. ID NO. 22 - TetB sense

- (pBR322: 1067-1091)

-----20nt UAUAGx4----- 25nt  
UAUAG UAUAG UAUAG UAUAG GTA GAT GAC GAC CAT CAG GGA CAG C  
\* \* \* \* \*

SEQ. ID NO. 20 - CmB antisense

(pACYC184: 3768-3793)

----- 21nt CAUX7 ----- 26nt  
CAU CAU CAU CAU CAU CAU CAU CAG GCG TTT AAG GGC ACC AAT AAC TG  
\* \* \* \* \*

SEQ. ID NO. 27 - CmA sense

(also designated 4555) - (pACYC184: 240-216)

-----20nt AUAUGx4----- 25nt  
AUAUG AUAUG AUAUG AUAUG TCA GGA GCT AAG GAA GCT AAA ATG G  
\* \* \* \* \*

SEQ. ID NO. 28 - Cm A antisense

(also designated 4557) - (pACYC184: 22-46)

-----20nt AUUAGx4----- 25nt  
AUUAG AUUAG AUUAG AUUAG CAG GCG GGC AAG AAT GTG AAT AAA G  
\* \* \* \* \*

SEQ. ID NO. 23 - Cm B sense

(pACYC184: 4086-4062)

-----20nt CUAAX4----- 25nt  
CUAAU CUAAX CUAAX CUAAX TAC GGT GAA AAC CTG GCC TAT TTC C  
\* \* \* \* \*

SEQ. ID NO. 11 - Amp antisense

pBR322 (4159-4136)

-----21nt UAGx7 ----- 24 nt



UAG UAG UAG UAG UAG UAG UAG UAG AAG AGT ATG AGT ATT CAA CAT TTC  
 \* \* \* \* \*  
 SEQ. ID NO. - 24 Amp sense

pBR322 (2460-2483)

-----21nt AUGx7----- 24 nt  
 AUG AUG AUG AUG AUG AUG AUG AAC GCA GGA AAG AAC ATG TGA GCA  
 \* \* \* \* \* \* \* \*

# *PCR reactions*

reaction	Tet A21	Tet B21	Cm B21	Cm A21
20µl primer 1	SEQ. ID 19	SEQ. ID 22	SEQ. ID 20	SEQ. ID 28
20µl primer 2	SEQ. ID 21	SEQ. ID 5	SEQ. ID 23	SEQ. ID 27

(primer concentration: 0.1 µg/µl)

## expected

fragment size:	977bp	423bp	357bp	257bp
reaction		Tet 21	Cm21	Amp21
20µl primer 1		SEQ. ID 19	SEQ. ID 20	SEQ. ID 11
20µl primer 2		SEQ. ID 5	SEQ. ID 27	SEQ. ID 24

(primer concentration: 0.1 µg/µl)

## expected

fragment size:	1489bp	756bp	1742bp
mix		x128	
2µl dNTP's (2.5 mM each)		256µl	
2.5µl 10xBuff.		320µl	
18µl H2O		2304µl	
0.1µl Taq (5 u/µl)		12.8µl	

put 360µl mix in each tube. add primers

add 1µl DNA (40 ng/µl)

Tet A21	Tet B21	Cm B21	Cm A21
pBR322	pBR322	pACYC184	pACYC84
Tet 21	Cm21	Amp21	
pBR322	pACYC184	pBR322	

## *Temperature regime*

94°C 40sec  
 40°C 2min  
 72°C 4min  
 30 cycles

72°C 5min

6°C infinitely

**[0080]** The required plasmids are illustrated in Figs. 1 through 6. Figures 1-3 depict plasmids constructed from fragments containing 12 nucleotide (nt) overhangs. Figures 4-6 depicts plasmids constructed from fragments containing 21 nt overhangs.

#### B. Connecting the fragments

**[0081]** The PCR fragments were mixed and their 3' overhangs exposed by utilizing UDG. The various reactions are given below. Two identical independent experiments were conducted. The DNA fragments for each experiment were prepared independently as well.

**[0082]** The concentration of the DNA fragments was 0.05 pmol/μl.

UDG Buff. x10 = 200mM Tris HCl pH 8.4, 500mM KCl, 15mM MgCl<sub>2</sub>.

**[0083]** Reactions mixtures:

	1 μl	1 μl	1μl	1μl	7μl	1.2μl	0.8μl
1. Amp12	Tet12	Cm12	H <sub>2</sub> O	H <sub>2</sub> O	UDG Buff.x10	UDG	
	1μl	1μl	1μl	1μl	7μl	1.2μl	0.8μl
2. Amp12	Tet12	CmA12	CmB12	H <sub>2</sub> O	UDG Buff.x10	UDG	
	1μl	1μl	1μl	1μl	7μl	1.2μl	0.8μl
3. Amp12	TetA12	TetB12	Cm12	H <sub>2</sub> O	UDG Buffx10	UDG	
	1μl	1μl	1μl	1μl	7μl	1.2μl	0.8μl
4. Amp21	Tet21	Cm21H <sub>2</sub> O	H <sub>2</sub> O	UDG Buff.x10	UDG		
	1μl	1μl	1μl	1μl	7μl	1.2μl	0.8μl
5. Amp21	Tet21	CmA21	CmB21	H <sub>2</sub> O	UDG Buff.x10	UDG	
	1μl	1μl	1μl	1μl	7μl	1.2μl	0.8μl
6. Amp21	TetA21	TetB21	Cm21	H <sub>2</sub> O	UDG Buff.x10	UDG	

**[0084]** The reactions were incubated at 37°C for 5 hours. 1 μl of each reaction was used in the electroporation of 20μl GibcoBRL ElectroMax DH10B cells. After the transformation the cells of reactions 1 and 4 were plated on LB plates containing Ampicillin, Tetracycline and Chloramphenicol, for selection (Figs. 8A and 8B respectively). Likewise, the cells of reactions 2 and 5 were plated on LB plates containing Ampicillin and Tetracycline (Figs. 8C and 8D respectively). Cells of reactions 3 and 6 were plated on LB plates containing Ampicillin and Chloramphenicol (Figs. 8E and 8F respectively). The fusion fragments CmA-CmB and TetA-TetB contain only the upstream and downstream fragments

of Cm<sup>r</sup> and Tet<sup>r</sup>. In addition they are interconnected by short linker repeats. They cannot therefore confer resistance to chloramphenicol or tetracycline.

**[0085]** An illustration of the constructs along with the number of transformants obtained, is given in Fig. 9. In the figure, colonies marked with a \* were checked by PCR as detailed below and proven to contain plasmids with rearrangements rather than the required constructs.

#### *Further examination of the constructs by PCR*

**[0086]** The 6 transformants of reactions 2 as well as 16 representatives from reaction 5 were PCR tested in order to check whether they contain the desired constructs: Since these transformants were resistant to both Ampicillin and Tetracycline it was assumed that these fragments had fused correctly to each other. It was desired to know whether the Cm A and Cm B were present as well. Therefore the existence of the Cm fragments was tested by checking the various transformants by PCR with primers SEQ. ID NO. 14 and SEQ. ID NO. 25. Since in this constructs, Cm A (which encompasses part of the 5' region of Cm) is 257bp long and Cm B (which encompasses part of the 3' region of Cm) is 357bp long, the overall size of the expected PCR product should be 614bp long, some 142bp shorter than the size of a complete Cm fragment (756bp long).

**[0087]** In Fig. 10(a), the six left lanes are of the transformants obtained from reaction 2, the next five lanes are from reaction 5. The right lane is a full Cm-fragment control.

**[0088]** In Fig. 10 (b), the right lane is a full Cm-fragment control. The rest of the lanes are of more transformants from reaction 5.

**[0089]** The results indicate that none of the transformants in reaction 2 contain the desired constructs. In reaction 5, 9 out of 16 transformants contain the desired product.

**[0090]** Transformants from reactions 4, 5 and 6 were further tested by PCR to check if they contain the desired fragments in the desired order, as detailed below:

reaction:	A	B	C	D
	Tet	Cm A Tet A	Cm B Amp	Amp Tet B
primer 1	SEQ. ID 13	SEQ. ID 26	SEQ. ID 17	SEQ. ID 16
primer 2	SEQ. ID 2	SEQ. ID 15	SEQ. ID 8	SEQ. ID 18
expected				
fragment size:	1473bp	1180bp	2043bp	1207bp

**[0091]** The tests carried out according to the above, verified that the transformants from reactions 5 and 6 (which were plasmids constructed from fragments having 21 nt overhangs)

contained the desired fragments in the correct order. In reaction 5, 9 out of 16 transformants contained the desired product. In reaction 6, 2 out of 2 transformants contained the desired product.

**[0092]** This experiment clearly demonstrates the advantage of constructing plasmids from fragments with long overhangs (around 21 nt). Almost no colonies were obtained in the 3 experiments (1-3) that used fragments with overhangs of 12 nt. The only 6 colonies that were observed did not contain the correct construct. In contrast, in the experiments that used fragments with overhangs of 21 nt, colonies containing the correct product were readily obtained.

#### *Example 4*

##### *Assembly of three DNA fragments using ExoIII-created overhangs*

**[0093]** Exonuclease III (ExoIII) of *E. coli* was used in order to create complementary overhangs of 20 bases or longer. ExoIII digests one strand of blunt-ended DNA or DNA containing 5' overhangs, in a 3' to 5' direction, thus creating a 5' overhang or enlarging an existing 5' overhang. By regulating the temperature and time of the reaction one can control the extent of the digestion, hence the length of the overhang.

**[0094]** The formation of a construct made of three fragments is illustrated. Using the same method, plasmids made out of four and five segments were also constructed (not shown).

**[0095]** The plasmid illustrated in Fig. 12 was assembled by the joining of three independently produced DNA fragments. These three DNA fragments are:

- a) a DNA fragment of 1739 base pairs (bp) containing the Amp<sup>r</sup> gene and the ColE1-ori region;
- b) a DNA fragment of 1466 bp containing the Tet<sup>r</sup> gene; and
- c) a DNA fragment of 745 bp containing the Cm<sup>r</sup> gene.

These fragments are illustrated schematically in Fig. 12 showing their relative positioning one to the other, namely, that the Cm<sup>r</sup> fragment was to be connected at its one end to one end of the Tet<sup>r</sup> fragment and at its other end to one end of the Amp<sup>r</sup> + ColE1-ori fragment, and likewise, the other ends of the Tet<sup>r</sup> and Amp<sup>r</sup> + ColE1-ori fragments were to be connected to each other to provide for a circular DNA molecule, being the desired plasmid, having the above predetermined order of the three fragments.

**[0096]** To produce the above three DNA fragments, the aforementioned PCR procedure was carried out using the following primers and template DNA:

a) The 1739 bp Amp<sup>r</sup> + ColE1-ori DNA fragment was synthesized by the PCR procedure using primers SEQ. ID NO. 29 (also designated 4142) and SEQ. ID NO. 30 (also designated 3884) and pBR322 as the template DNA. The concentrations of the primers and template DNA, as well as the other PCR conditions, are as indicated above. The relative direction of the primers with respect to the synthesis of the Amp<sup>r</sup> + ColE1-ori fragment as it is positioned in the completed plasmid product is as depicted schematically in Fig. 12. Thus, primer SEQ. ID NO. 29 was synthesized to have a predetermined sequence so as to provide for the desired junction region between the Amp<sup>r</sup> + ColE1-ori fragment and the Cm<sup>r</sup> fragment, and primer SEQ. ID NO. 30 was synthesized to have a predetermined sequence so as to provide for the desired junction between the Amp<sup>r</sup> + ColE1-ORI fragment and the Tet<sup>r</sup> fragment.

b) The 1466 bp Tet<sup>r</sup> DNA fragment was synthesized by the PCR procedure using primers SEQ. ID NO. 31 (also designated 31160) and SEQ. ID NO. 32 (also designated 30397) and pBR322 as the template DNA. The concentrations of the primers and template DNA, as well as the other PCR conditions, are as indicated above. The relative direction of the primers with respect to the synthesis of the Tet<sup>r</sup> fragment as it is positioned in the completed plasmid product is as depicted schematically in Fig. 12.

c) The 745 bp Cm<sup>r</sup> DNA fragment was synthesized by the PCR procedure using primers SEQ. ID NO. 33 (also designated 3595) and SEQ. ID NO. 34 (also designated 4143) and pACYC184 as the template DNA. The relative direction of the primers with respect to the synthesis of the Cm<sup>r</sup> fragment as it is positioned in the completed plasmid product is as depicted schematically in Fig. 12.

**[0097]** The sequences of the above primers are as follows:

The arrows inside the sequences mark the junction points.

primers SEQ. ID NO. 29 and SEQ. ID NO. 30

for the Amp<sup>r</sup> + ColE1-ori segment:

	part of Cm <sup>r</sup> region sequence	part of Amp <sup>r</sup> region sequence
primer SEQ. ID NO. 29:	ATTGGTGGCCTTAACGCCTG↓	AACGCAGGAAGAACATGTG
	(also designated 4142)	

	part of Tet <sup>r</sup> region sequence	part of Amp <sup>r</sup> region sequence
primer SEQ. ID NO. 30:	AGCGTTGGGTCCTGG↓	CCAAAGAGTATGAGTATTCAACA
	(also designated 3884)	

primers SEQ. ID NO. 31 and SEQ. ID NO. 32

for the Tet<sup>r</sup> segment:

part of  
Cm<sup>r</sup> region sequence
part of Tet<sup>r</sup> region sequence  
 primer SEQ. ID NO. 31: AGCTCCTGA↓TTCTCATGTTTGACAGCTTATC  
 (also designated 31160)

part of  
Amp<sup>r</sup> + ColE1-ori
part of Tet<sup>r</sup> region sequence  
 primer SEQ. ID NO. 32: ATACTCTT↓TGGCCAGGACCCAACGCTGCC  
 (also designated 30397)

primers SEQ. ID NO. 33 and SEQ. ID NO. 34

for the Cm<sup>r</sup> segment:

part of  
Tet<sup>r</sup> region
part of Cm<sup>r</sup> region sequence  
 primer SEQ. ID NO. 33: AAACATGAGAA↓TCAGGAGCTAAGGAAGCTAAATG  
 (also designated 3595)

The Amp region
The CM region  
 primer SEQ. ID NO. 34: ATGTTCTTTCTGCGTT↓CAGGCGTTTAAGGGACCAATAAC  
 (also designated 4143)

**[0098]** In view of the fact that the original primer sequences were derived only from the Tet<sup>r</sup>, Cm<sup>r</sup> and Amp<sup>r</sup> + ColE1-ori region genes, with no introduction of any additional ("linker") DNA sequences whatsoever, the above junctions between said genes are "seamless".

**[0099]** The PCR was carried out according to the description in the former examples. Each PCR fragment, separately, was then subjected to agarose-gel purification using a commercial kit, namely, Bio-Rad's Prep-A-Gene™ purification kit and following the manufacturer's instructions.

**[0100]** Once purified, each PCR fragment was then quantitated by determining the DNA concentration of each fragment by standard procedures, using Pharmacia's Gene-Quant™ RNA/DNA calculator and following the manufacturer's instructions.

**[0101]** The PCR fragments were then subjected to Exo III digestion and subsequent joining procedure. The various PCR fragments were mixed together (0.15 pmol DNA for each fragment), in a cooled (6°C) reaction mixture of 12 µl containing: 1.2 µl 10 x TA buffer (330 mM Tris-acetate, pH 7.8; 660 mM potassium acetate, 100 mM magnesium acetate and 5 mM DTT); 0.8 µl Exo III (200 U/µl purchased from Epicentre Technologies); and sterile

double distilled H<sub>2</sub>O to make up the final volume of 12  $\mu$ l. In practice, the PCR fragments were mixed into a precooled, namely, 6°C, TA buffer solution made up to 11.2  $\mu$ l with the H<sub>2</sub>O, to which was then added the 0.8  $\mu$ l Exo III. Adding the Exo III last provides for better control over the Exo III reaction, which reaction is controlled by the time of incubation of the PCR fragments with the Exo III. Following the Exo III addition, the reaction mixture was then incubated at 6°C for 40 mins. (the time necessary to achieve more than 20 nucleotide degradation of each DNA strand in the 3'-5' direction under the above conditions of temperature = 6°C and concentration of Exo III).

**[0102]** The Exo III reaction was then stopped by performing a phenol extraction. This is done by adding to the above reaction mixture an equal volume = 12  $\mu$ l of a 1:1 (v/v) phenol/chloroform mixture which causes denaturation of the Exo III. The aqueous phase was then separated from the above phenol/chloroform mixture, this aqueous phase containing the PCR fragments. The separated aqueous phase was then subjected to three washes with chloroform to yield a final aqueous phase having an essentially purified mixture of the Exo III-digested PCR fragments.

**[0103]** In order to prevent both the evaporation of the buffer and the drying of the fragments, 40  $\mu$ l of mineral oil was added to the vessel containing the ExoIII-digested PCR fragments. The mixture was then heated to 75°C, at which temperature it was further incubated for one hour. After this incubation, the mixture is slowly cooled, under conditions providing only a 2°C decrease in temperature per hour, until it reached 37°C (the heating and cooling to provide for specific complementary interactions between complementary overhangs on the PCR fragments and to prevent non-specific interactions). This heating and cooling represents the first stage of the specific joining between the fragments, the joining by way of hydrogen bonding between complementary overhangs.

**[0104]** Once the above mixture has reached 37°C, it was then subjected to the final stage of the joining, including the filling in reaction, as follows:

**[0105]** To the cooled (at 37°C) mixture of now essentially joined (by hydrogen bonding of complementary overhangs) PCR fragments there was added 10  $\mu$ l of the 'Synthesis Mixture', which contains:

1  $\mu$ l of 20 mM ATP

4  $\mu$ l of 2.5 mM dNTPs (= dATP, dTTP, dCTP and dGTP in equal amounts, concentration of each = 2.5 mM)

1  $\mu$ l of 10 x TA buffer (see above for constituents)

1  $\mu$ l of T7 DNA polymerase (5 U/ $\mu$ l, purchased from USB)

1  $\mu$ l of T4 DNA ligase (10 U/ $\mu$ l, purchased from Epicentre Technologies)

2  $\mu$ l of sterile double-distilled H<sub>2</sub>O

Total volume: 10  $\mu$ l

**[0106]** The above synthesis mixture had the T7 DNA polymerase and dNTPs to facilitate the filling in of gaps in the junction regions between the joined fragments, as well as the ATP and the T4 ligase to covalently join the DNA strands together once the filling in of gaps has been completed.

**[0107]** Following the addition of the synthesis mixture, the resulting reaction mixture was then incubated for two hours at 37°C. After this incubation, the reaction mixture was then ethanol precipitated under standard conditions to finally yield a pellet of precipitated DNA which was essentially the completed DNA construct composed of the joined PCR fragments. This DNA pellet was resuspended in 5  $\mu$ l sterile double-distilled H<sub>2</sub>O and is ready for further analysis or use.

**[0108]** Following the preparation of the desired construct described above from the three PCR fragments, this construct was analyzed for its biological activity, namely, whether or not it could confer resistance to all three antibiotics when introduced into bacterial cells. Thus, electroporation of electrocompetent DH10B *E. coli* cells was performed using a 2  $\mu$ l aliquot of a 5  $\mu$ l final product containing the DNA construct. After electroporation, the cells were first plated on agar plates containing ampicillin. The results revealed more than 1000 colonies on these plates, indicating that more than 1000 originally transformed cells received a DNA construct having at least an active Amp<sup>r</sup> gene. Of these Amp<sup>r</sup> colonies, 40 were chosen at random, as a test sample, and were plated on both tetracycline- and chloramphenicol-containing agar plates. All 40 of these test colonies grew on these plates as well, indicating that they were also Tet<sup>r</sup> and Cm<sup>r</sup>. Hence, it is concluded that at least these 40 colonies received an intact construct in which all of the Tet<sup>r</sup>, Cm<sup>r</sup> and Amp<sup>r</sup> genes were intact and fully expressible. Some of the colonies were tested further by PCR, as indicated above. The PCR bands that appeared were of the expected sizes.

**[0109]** In the above example, the ligation and fill-in steps have been used for the sake of completeness and to illustrate possible alternative procedures. However, as will be easily understood by the skilled person, said ligation and fill-in steps are not necessary, and the procedure exemplified above can be carried out without using such steps. Examples of such procedures without said steps have not been given, for the sake of brevity.



### Example 5

#### *Assembly of 8 DNA fragments into a circularized plasmid using UDG-created overhangs*

**[0110]** The plasmid to be constructed was designed to consist of eight fragments, each to be prepared separately by PCR amplification and then joined in a specific directional fashion to provide a circularized plasmid as the end-product. For the purposes of exemplifying the method of the present invention, it was chosen to combine eight PCR fragments which together encompass four regions, as detailed below, thus carrying out a specific directional connection of eight independent PCR fragments to form a single active plasmid construct, and this by an essentially one-step procedure in accordance with the present invention, a result never obtained or attempted in the prior art.

**[0111]** This example is similar to Example 1, as the overhangs are created by the use of dU in the primer and the enzyme UDG to expose the overhang. However, in the present example, only one dU residue per primer is used. Therefore, the procedure of exposing the overhang comprises an additional reagent, N,N Dimethyl-ethylenediamine. The use of this reagent is explained in further detail in the section of general procedures above and in the above mentioned copending patent application, identified as Attorney Docket 4149/96, Israeli Patent Application No. 120337. In addition, the present example uses primers that contain, at their 5' end, the natural gene sequence of the fragment to which they are to be joined. The junction is therefore seamless: no unnecessary residues are added at the junctions. In contrast, Example 1 uses primers that contain at their 5' end an irrelevant sequence whose only purpose is to serve as an overhang, since several dU residues must be used in the 5' part of the primer sequence. This irrelevant sequence is then introduced into the resulting fragment between the natural gene sequences of the fragments that are joined. Thus, the use of only one dU nucleotide, as in the present example, enables the seamless joining of fragments.

**[0112]** In Fig. 13, there is shown schematically the plasmid that was designed and produced by the method of the present invention. This plasmid carries four independent antibiotic resistance genes, for resistance to Ampicillin (Amp<sup>r</sup> gene, or hereinafter Amp); Tetracycline (Tet<sup>r</sup> gene, or hereinafter Tet); Chloramphenicol (Cm<sup>r</sup> gene, or hereinafter Cm) and Kanamycin (Kn<sup>r</sup> gene, or hereinafter Kn). The plasmid also carries the ColEI origin of replication (ColEI-ori), which in this specific instance is situated next to the Amp<sup>r</sup> gene.

Hence, such a plasmid is capable of being replicated in a host cell and will endow the host cell with resistance to all four types of antibiotic.

**[0113]** Accordingly, it is possible to readily select for those host cells transformed by this plasmid by growing the transform cells in the presence of one of the antibiotics and then to screen for the resistance to the other antibiotics. In order to verify that mis-connections did not occur, additional verification tests may be carried out. One such test carried out included testing transformed colonies by restriction enzyme analysis. Plasmid DNA was prepared from a number of transformed colonies and checked by restriction enzymes.

**[0114]** For this eight fragments construction, the following was carried out:

(i) *Preparation of the specific primers for the PCR amplification.*

**[0115]** As shown in Fig. 13, and as outlined above, it was desired to construct a circularized plasmid having four regions:

- (a) a Tet<sup>r</sup> region.
- (b) an Amp<sup>r</sup> + ColEI-ori region.
- (c) a Cm<sup>r</sup> region.
- (d) a Kn<sup>r</sup> region.

**[0116]** The order of the connection is depicted in figure 13.

**[0117]** To achieve this task, eight separate fragments were designed:

Name of fragment		Size(bp)	plasmid	Location
AmpA	241365 3885	574	pBR322	3603-4159
AmpB	27342 241386	1171	pBR322	2480-3624
CmA	40122 27343	481	PACYC18	4021-240
CmB	25596 4144	293	PACYC18	3768-4043
TetA	27341 36176	862	pBR322	1-848
TetB	30402 25595	631	pBR322	827-1449
KnA	31254 25953	569	PACYC17	1809-2368
KnB	25952 31253	652	PACYC17	2349-2991

**[0118]** Written are the names of the fragments, their sizes, the plasmid from which they were PCR amplified, the exact location on the original plasmid of the site to be amplified and the numbers of the primers that were used to amplify the fragments. The full sequence and maps of the various regions of these plasmids are well known and can be accessed from GenBank database under accession nos. J01749 (pBR322), X06403 (pACYCI84) and X06402 (pACYC177).

**[0119]** The AmpB fragment includes the 5' part of the Amp fragment and the ColEI-ori sequence. The AmpA fragment includes the 3' part of the Amp fragment. The CmA fragment includes the 5' part of the Cm fragment and the CmB fragment includes the 3' part of the Cm fragment. The TetB fragment includes the 5' part of the Tet fragment and the TetA fragment includes the 3' part of the Tet fragment. The KnA fragment includes the 5' part of the Kn fragment and the KnB fragment includes the 3' part of the Kn fragment.

**[0120]** Each primer consists of two regions: a 3' region complementary to the DNA to be amplified, and a 5' region complementary to the fragment it should be connected to. Using standard automated procedures to produce polynucleotide oligomers (Applied Biosystems, U.S.A.), the following primers were synthesized:

*Primers for the amplification of the AmpA fragments*

**[0121]** primer SEQ. ID NO. 35, also designated 241365:  
internal Amp<sup>r</sup> region

ATTGCTGCAGGCATCGTGGTGUCA  
\*

primer SEQ. ID NO. 36, also designated. 3885:

Cm <sup>r</sup> region	Amp <sup>r</sup> region
AGCGTTGGGTCTCGCCA - AAGAGTAUGAGTATTCAA *	

*Primers for the amplification of the AmpB fragments*

**[0122]** primer SEQ. ID NO. 37, also designated 27342:

Cm <sup>r</sup> region	Amp <sup>r</sup> region
ACGCCTG - AACGCAGAAAGAACAUGTG *	

primer SEQ. ID NO. 38, also designated 241366:

internal Amp<sup>r</sup> region  
ACACCACGATGCCTGCAGCAAUGG

\*

*Primers for the amplification of the CmA fragments*

**[0123]** primer SEQ. ID NO. 39, also designated 40122:

Kn<sup>r</sup> region Cm<sup>r</sup> region

AGG CCT GGT ATG AGT C - TCA GGA GCU AAG GAA GCT AAA ATG

\*

primer SEQ. ID NO. 40, also designated 27343:

internal Cm<sup>r</sup> region

ATTGGCTGAGACGAAAAACATAUTCTC

\*

*Primers for the amplification of the CmB fragment*

**[0124]** primer SEQ. ID NO. 41, also designated 25596:

internal Cm<sup>r</sup> region

ATATGTTTTTCGTCTCAGCCAAUCC

\*

primer SEQ. ID NO. 42, also designated 4144:

Amp<sup>r</sup> region Cm<sup>r</sup> region

ATGTTCTTTCCTGCGTT - CAGGCGUTTAAGGGCACCAATAAC

\*

*Primers for the amplification of the TetA fragment:*

**[0125]** primer SEQ. ID NO. 43, also designated 27341:

internal Tet<sup>r</sup> region

ATACCGCAAGCGACAGGCCGAUCATCG

\*

primer SEQ. ID NO. 44, also designated 36176:

Kn antisense Tet sense

ACGTGGCTTTGTTG - TTCTCATGTGTGACAGCTTATC

\*

*Primers for the amplification of the TetB fragment:*

**[0126]** primer SEQ. ID NO. 45, also designated 30402:

Amp<sup>r</sup> region Tet<sup>r</sup> region

ATACTCTT - TGGCCAGGACCAACGCGCCCC

\*

primer SEQ. ID NO. 46, also designated 25595:

internal Tet<sup>r</sup> region

ATCGGCCTGTCGCTTGC GGTAUTCG

\*

*Primers for the amplification of the KnA fragment:*

**[0127]** primer SEQ. ID NO. 47, also designated 31254:

Tet region

Kn region

ACATGAGAA - CAACAAAGCCACGUTGTGTCTC

\*

primer SEQ. ID NO. 48, also designated 25953:

internal Kn region

AGACGAAATACGCGATCGCUGTTAA

\*

*Primers for the amplification of the KnB fragment:*

**[0128]** primer SEQ. ID NO. 49, also designated 25952:

internal Kn region

AGCGATCGCGTATTTCGTCTCUCGCTC

\*

primer SEQ. ID NO. 50, also designated 31253:

Cm region

Kn region

AGCTCCTGA - GACTCATACCGGCCUGAATCG

\*

*(ii) Preparation of the PCR fragments*

**[0129]** PCR reactions were carried out as in the other examples above. Following PCR synthesis of the individual fragments, each fragment was purified by standard agarose-gel purification techniques using the commercially available Bio-Rad "Prep-A-Gene<sup>TM</sup>" DNA purification kit and adhering to the manufacturer's instructions. Following purification, the concentration of the purified fragment DNA was determined by standard procedures using the Pharmacia "Gene-Quant<sup>TM</sup>" RNA/DNA Calculator" and adhering to the manufacturer's instructions.

*(iii) Connection of the PCR-produced fragments*

**[0130]** The eight PCR fragments as produced and purified according to the above-mentioned procedure, were connected to each other in a one-step reaction mixture in a single reaction vessel. This was achieved by mixing the fragments together in a 25µl reaction

mixture that included: 0.15 pmol of each fragment, 2.5µl buffer (200mM Tris-HCl pH 8.4, 500 mM KCl, 15mM MgCl<sub>2</sub>), 2.5 µl of 1M N,N-Dimethyl-ethylenediamine and 6.25 units of UDG (GibcoBRL). The mixture was incubated at 37°C for 4 hours and then transferred to 70°C for 5 minutes to facilitate dissociation of the short nicked single-stranded DNA from the 5' ends of the fragments. After dissociation the short nicked single-stranded DNA were removed using "QIAquick PCR purification kit" (QIAGEN) adhering to the manufacturer's instructions. Before adding the first buffer of the kit, 200 µl of hot (70°C) buffer (20mM Tris-HCl pH 8.4, 50 mM KCl, 1.5mM MgCl<sub>2</sub>) was added in order to minimize the reannealing of the short nicked single-stranded DNA. The DNA was eluted in 30µl of double-distilled water (DDW). 27µl of the DNA was incubated with 3µl of buffer (20mM Tris-HCl pH 8.4, 50 mM KCl, 1.5mM MgCl<sub>2</sub>) in a water-bath, at 70°C. The bath was shut down and the temperature was slowly decreased to 37°C. This allows for the joining between the complementary overhangs, while minimizing illegitimate connection.

*(iv) Analysis of the efficiency of joining of the fragments*

**[0131]** A 1 µl sample of the above DNA, containing the newly constructed plasmid, made from joining the 8 separate fragments, was used to transform *E. coli* DH10B cells by electroporation. In the transformation procedure 20 µl of electrocompetent "ElectroMax" cells (purchased from Gibco BRL) were mixed with the above DNA sample and subjected to electroporation in a commercially available apparatus (BioRad "*E. coli* Pulser Apparatus" set at 1.8 kV and operated according to the manufacturer's instructions).

**[0132]** Following electroporation (transformation) the cells were plated on LB Agar plate containing 100 mM Ampicillin (to select for transformants having Ampicillin resistance by virtue of having being transformed with a DNA carrying the Amp<sup>r</sup> gene). On the day after, three colonies were picked and checked for resistance to Chloramphenicol, Tetracycline and Kanamycin by plating on Agar plates containing the appropriate antibiotics.

**[0133]** The results showed that the three colonies were resistant to all the four antibiotics. The colonies were further checked by restriction enzyme analysis and proved to be correct (data not shown).

**[0134]** The above result is very significant since it shows that it is possible to join correctly 8 separate DNA fragments in a specific directional manner, in a single reaction mixture by an essentially one-step procedure.

**[0135]** All the above description of preferred embodiments and examples have been provided for the purpose of illustration, and are not meant to limit the invention. Many modifications can be made in the methods, materials and conditions, and many different products and results can be obtained, using different numbers of overhangs and different numbers of fragments to be joined, all without exceeding the scope of the invention.

## SEQUENCE LISTING

### (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) NAME: Gil Sharon and Gesher Israel Advance Biotechs (1996) Ltd.
  - (B) STREET: POB 98
  - (C) CITY: Beit Neqofa
  - (E) COUNTRY: Israel
  - (F) POSTAL CODE (ZIP): 90830
  - (G) TELEPHONE: +972 2 5700 381
  - (H) TELEFAX: +972 2 5700 586
- (ii) TITLE OF INVENTION: Improved DNA Assembly Method
- (iii) NUMBER OF SEQUENCES: 50
- (iv) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

### (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AGCTCCTGAT TCTCATGTTT GACAGCTTAT C

31

### (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CUACUACUAC UATGCCAGG ACCCAACGCT GCCCGAG

37

### (2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 40 base pairs



- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CUACUACUAC UACUATGGCC AGGACCCAC GCTGCCGAG

40

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 43 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CUACUACUAC UACUACUATG GCCAGACCC AACGCTGCCC GAG

43

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 46 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CUACUACUAC UACUACUACU ATGGCCAGGA CCCAACGCTG CCCGAG

46

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 55 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CUACUACUAC UACUACUACU ACUACUACUA TGGCCAGGAG CCAACGCTGC CCGAG

55

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 41 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATTGGTGCCC TTAAACGCCT GAACGCAGGA AAGAACATGT G

41

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

UAGUAGUAGU AGAAGAGTAT GAGTATTCAT CATTTC

36

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 39 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

UAGUAGUAGU AGUAGAAGAG TATGAGTATT CAACATTTC

39

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 42 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

UAGUAGUAGU AGUAGUAGAA GAGTATGAGT ATTCAACATT TC

42

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

UAGUAGUAGU AGUAGUAGUA GAAGAGTATG AGTATTCAAC ATTTC 45

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 54 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

UAGUAGUAGU AGUAGUAGUA GUAGUAGUAG AAGAGTATGA GTATTCAACA TTTC 54

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AUCAUAUCAU AUUCTCATG TTGACAGCT TATCATC 37

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CAUCAUCAUC AUCAGGCCTT TAAGGCACC AATAACTG 38

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

UACUAUACUA UACATGCCGG CGATAATGGC CTGCTTC 37

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

UAUAGUAUAG UAGTAGATGA CGACCATCAG GGACAGC 37

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AUCUAAUCUA AUTACGGTGA AAACCTGGCC TATTTC 37

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AUGAUGAUGA UGAACGCAGG AAAGAACATG TGAGCA 36

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CAUAUCAUUAU CAUAUCAUUAU TTCTCATGTT TGACAGCTTA TCATC 45

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CAUCAUCAUC AUCAUCAUCA UCAGGCGTTT AAGGGCACCA ATAACTG 47

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CUAUACUAUA CUAUACUAUA CATGCCGGCG ATAATGGCCT GCTTC 45

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

UUAUGUAUAG UUAUGUAUAG GTAGATGACG ACCATCAGGG ACAGC 45

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CUAAUCUAAU CUAUCUAAU TACGGTGAAA ACCTGGCCTA TTTC

45

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AUGAUGAUGA UGAUGAUGAU GAACGCAGGA AAGAACATGT GAGCA

45

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AUAUGAUAUG AUTCAGGAGC TAAGGAAGCT AAAATGG

37

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

AUUAGAUUAG AUCAGGCGGG CAAGAATGTG AATAAAG

37

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

AUAUGAUAUG AUAUGAUAUG TCAGGAGCTA AGGAAGCTAA AATGG

45

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

AUUAGAUUAG AUUAGAUUAG CAGCGGGCA AGAATGTGAA TAAAG

45

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 41 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATTGGTGCCC TTAACGCCT GAACGACGGA AAGAACATGT G

41

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

AGCGTTGGGT CCTGGCCAAA GAGTATGAGT ATTCAACA

38

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

AGCTCCTGAT TCTATGTTT GACAGCTTAT C

31

(2) INFORMATION FOR SEQ ID NO: 32:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ATATCTTTTG GCCAGGACCC AACGCTGCCC

30

(2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 35 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

AAACATGAGA ATCAGGAGCT AAGGAAGCTA AAATG

35

(2) INFORMATION FOR SEQ ID NO: 34:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 41 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

ATGTTCTTTTC CTGCGTTCAG GCGTTTAAGG GCACCAATAA C

41

(2) INFORMATION FOR SEQ ID NO: 35:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

ATTGCTGCAG GCATCGTGGT GUCA

24

(2) INFORMATION FOR SEQ ID NO: 36:



- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AGCGTTGGGT CCTGGCCAAA GAGTAUGAGT ATTCAA

36

(2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

ACGCCTGAAC GCAGGAAAGA ACAUGTG

27

(2) INFORMATION FOR SEQ ID NO: 38:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

ACACCACGAT GCCTGCAGCA AUGG

24

(2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGGCCTGGTA TGAGTCTCAG GAGCUAAGGA AGCTAAAATG

(2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

ATTGGCTGAG ACGAAAAACA TAUTCTC

27

(2) INFORMATION FOR SEQ ID NO: 41:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

ATATGTTTTT CGTCTCAGCC AAUCC

25

(2) INFORMATION FOR SEQ ID NO: 42:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 41 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42

ATGTTCTTTC CTGCGTTCAG GCGUTTAAGG GCACCAATAA C

41

(2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATACCGCAAG CGACAGGCCG AUCATCG

27

(2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

ACGTGGCTTT GTTGTCTCA TGUTTACAG CTTATC

36

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

ATACTCTTTG GCCAGGACCC AACGUGCCC

30

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

ATCGGCCTGT CGTTGCGGT AUTCG

25

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

ACATGAGAAC AACAAAGCCA CGUTGTGTCT C

31

(2) INFORMATION FOR SEQ ID NO: 48:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AGACGAAATA CGCGATCGCU GTTAA

25

(2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AGCGATCGCG TATTTCGTCU CGCTC

25

(2) INFORMATION FOR SEQ ID NO: 50:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

AGCTCCTGAG ACTCATACCA GGCCUGAATC G

31